Molecular Biology Select

Cell

The duplication of genes, of noncoding DNA, and occasionally of entire genomes has contributed to biological and genetic diversity. Recent studies examine both the benefits and challenges associated with duplication events and highlight some of the existing unsolved mysteries of molecular evolution.

Polyploid Chromosomes Meet Their Match

Wheat in flower. Photo courtesy of G. Moore.

Although the ancestral form of wheat is diploid, some commercial varieties are tetra-or even hexaploid. The existence of up to three related pairs of chromosomes (referred to as A, B, and D) poses a particular problem during meiosis because pairing of homologous chromosomes is a prerequisite for the first meiotic division. It is known that polyploid wheat is functionally diploid at meiosis, meaning that inappropriate pairing between related chromosomes (such as AB pairing) is prevented, whereas pairing between homologous chromosomes progresses normally (for example, AA pairing). Remarkably, a single locus in wheat on chromosome 5B, known as *Ph1,* suppresses pairing between related chromosomes. Graham Moore and colleagues (Griffiths et al.) have now narrowed the search for *Ph1* activity to a 2.5-megabase region. From their analysis, the most striking feature of this region—and a feature that clearly differentiates it from chromosomes 5A and 5D—is the presence of an insertion that originated from a subtelomeric region of chromosome 3A. This insertion contains tandem arrays of a 2.4 kb noncoding repeat sequence and interrupts a cluster of genes that encode proteins related to the protein kinase *cdc2*. Future work may establish whether the regulation of this *cdc2* cluster is altered by the insertion and whether this may affect meiotic pairing. An alternative possibility is that the noncoding tandem repeats by themselves promote correct pairing during meiosis. The pairing dilemma in polyploid wheat may not be unique. Similar strategies to ensure correct pairing in meiosis may have evolved in other organisms following genome duplication. Although the *Ph1* locus has broad effects on chromosome pairing, there could also be proteins and/or noncoding DNA that contribute to the pairing between specific chromosomes.

*S. Griffiths et al. (2006). Nature **439**, 749–752.*

Gene Duplicates, the More the Merrier?

Gene duplication may provide the raw material that enables establishment of new attributes by freeing one of the duplicate genes from the constraints of natural selection that would normally limit its ability to change. Surprisingly, Andrew Paterson and colleagues (Chapman et al.) report that genes in both *Arabidopsis* and rice that were duplicated as part of ancient genome duplication events show evidence of greater conservation than single-copy genes. The authors examined single nucleotide polymorphisms (SNPs) in the coding regions of duplicated and single-copy genes. Surprisingly, the SNPs in duplicated genes result in amino acid changes that are less dramatic than those in proteins encoded by single-copy genes. Moreover, gene duplicates have a higher fraction of SNPs in the 3rd position of the codon, which would be unlikely to result in an amino acid change. These results argue that genome duplication should not be viewed simply as a vehicle for gene diversification but as an event that can provide a spectrum of selective advantages to a species, which may include the buffering of crucial functions.

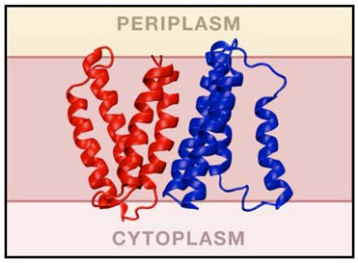
*B.A. Chapman et al. (2006). Proc. Natl. Acad. Sci. USA **103**, 2730–2735.*

Gene Duplication Puts Alternative Splicing on the Back Burner

Gene duplication is only one way that diversity can be generated. At least in higher eukaryotes, alternative splicing of messenger RNAs also plays an important part in the generation of diversity. Xun Gu and coworkers (Su et al.) report that in the human genome there is an inverse relationship between the size of a gene family and the number of alternative splice variants each gene encodes. This suggests that alternative splicing and gene duplication are closely linked means to generate protein diversity. Su et al. contend that the connection is strong enough that a decrease in the occurrence of alternative splicing may quickly follow (at least on an evolutionary time scale) the appearance of gene duplication. The analysis by Su et al. relies on the characterization of gene duplicates that have been created after the split between mouse and humans, which took place ~90 million years ago. Further temporal resolution of this evolutionary process may be possible if similar analyses are conducted for gene duplicates that have originated after a more recent split, for example between humans and chimpanzees. A question for future studies is what kind of selective pressures, if such exist, might favor the use of alternative splicing over duplication and vice versa.

*Z. Su et al. (2006). Genome Res. **16**, 182–189.*

Two-Timing Proteins Are Forced to Choose



Both proteins of the EmrE homodimer have identical amino acid sequences yet have opposite orientations in the membrane. Courtesy of G. von Heijne.

Following up on an extensive analysis of the topology of proteins in the inner membrane of the bacterium *Escherichia coli*, Gunnar von Heijne and colleagues (Rapp et al.) now describe a pattern of evolution for dual-topology proteins. These proteins, of which some are involved in multidrug resistance, are so named for their ability to insert into membranes in either orientation. Previous work from von Heijne and colleagues has shown that the orientation of inner-membrane proteins can be predicted by the distribution of the positively charged residues arginine (R) and lysine (K) in the loops between membrane segments. The side of the threaded protein with the greater positive charge typically orients toward the cytoplasmic side of the inner membrane. Consistent with this prediction, the dual-topology candidates all had (R+K) biases close to zero and, in most cases, could be made to adopt predictable orientations by altering the number of positively charged residues in the first loop. Rapp et al. then searched a variety of different bacterial genomes for homologs to the *E. coli* dual-topology proteins and grouped the genes into the ''pairs'' category when evidence of a duplicate gene was found; if no pair could be found they were put into the ''singletons'' category. Singletons consistently had a low charge bias (and should therefore also behave as dual-topology proteins). However, pairs had a much greater charge bias, such that one member of the pair favored the outside-in orientation and the other was oriented in the opposite way. Rapp et al. conclude that gene duplication allows each pair to become specialized for insertion into membranes in one particular orientation. This hypothesis is consistent with the recently published structure of a dual-topology protein EmrE, which forms a homodimer with each individual protein in the opposite orientation. However, it is not certain whether these specialized pairs of heterodimers provide a selective advantage over homodimers of dual-topology singletons because of enhanced function, or whether they are retained simply because once specialization occurs, it is impossible to remove one of the pair without affecting fitness.

M. Rapp et al. (2006). Nat. Struct. Mol. Biol. ***13****, 112–116.*

Noncoding Repeats Come into View

Unlike the proteins encoded by duplicated genes, for which functions can often be deduced by analyzing the homologous proteins of other species, less is known about duplicated noncoding segments. Eric Lander and colleagues (Kamal et al.) now report the discovery of an abundant and conserved family of mammalian repeat elements, called MER121. Kamal et al. show that in the human genome, there are more than 900 representatives of the MER121 family, which has a consensus sequence with a length of 412 base pairs. These elements are preferentially found in regions lacking genes and there is no evidence that MER121 sequences themselves code for either RNA or protein. Instead, the authors speculate that MER121 sequences have a structural or regulatory role and propose that an ancient transposable element dispersed the original sequence throughout the genome after the split in the evolutionary tree between mammals and birds. Analysis of sequence alignments of the MER121 family may provide clues that will ultimately lead to an understanding of the function of this mysterious repeat. Although individual bases are not conserved in the alignment, significant trends emerge if it is examined in larger segments of six bases. Curiously, from this analysis, the peak of highest conservation corresponds to a palindromic sequence, CATATG. Palindromes are commonly used in the recognition of DNA by proteins. Whether MER121 is a target of specific DNA binding proteins is an intriguing possibility that awaits further study.

M. Kamal et al. (2006). PNAS ***103****, 2740–2745.*

Robert P. Kruger